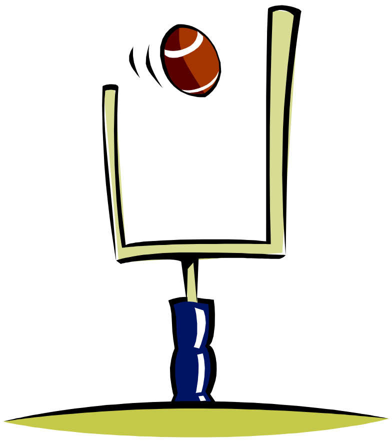
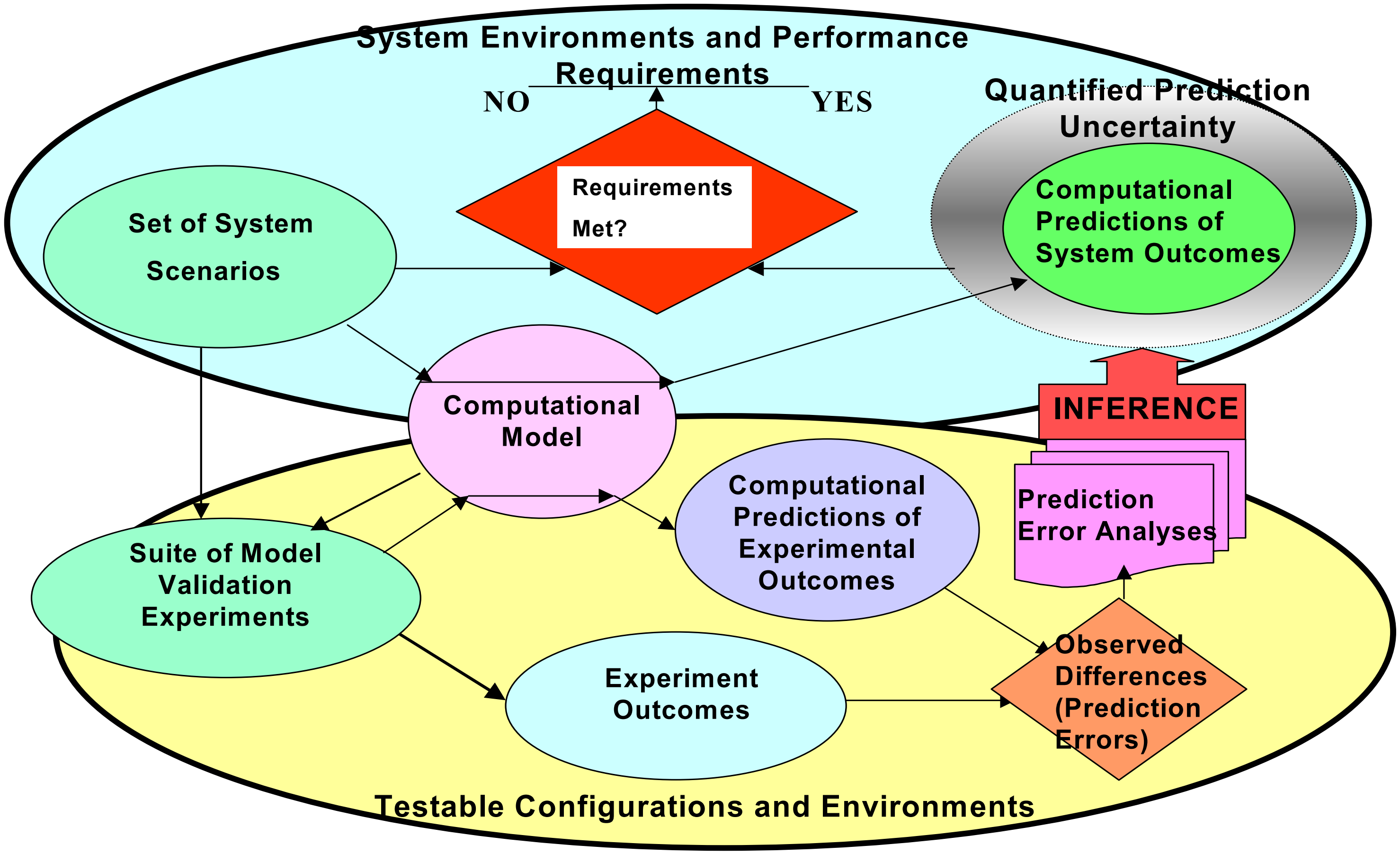
Extra Points

- Process schematic -- discussion
- Roles for computational modeling
- φ-estimation error
- Distribution prediction
- Comments
- Model-val as hypothesis testing
- UQ vis a vis model-val
- Issue: too much testing required

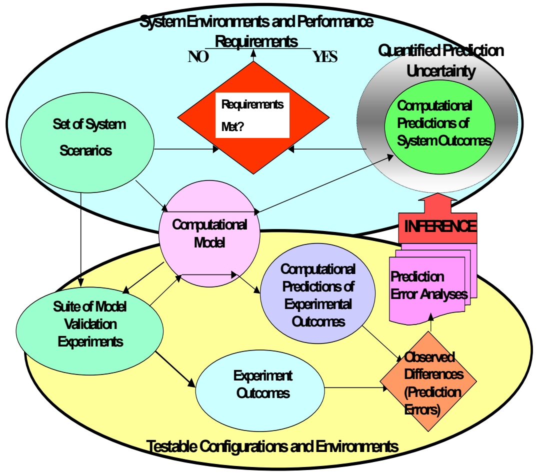


Measuring Predictive Capability: Purpose and Process



Evolving Views of this Schematic

1. Depicted my understanding of what people wanted to accomplish with "validated" computational models
2. My view: If you're serious about model-validation, here's what is required
3. Illustrates why:
   - Modeling has not achieved the supremacy claimed for it
     - » *Model-based certification is perhaps an unrealistic expectation*
     - » *You can do a lot of work in bottom ellipse and still not bridge the gap to applications*
   - Validation is regarded as a burden



• WANTED: realistic expectations

Some Thoughts on Computational Modeling

(adapted from presentation by Ernie Seglie, Science Advisor, DoD OT&E)

- Oversold
  - Replacement for testing
  - Decision agent

- More realistic expectations for modeling
  - Hypothesis generation
  - Scenario generation
  - Guide in an iterative "rolling assessment" of performance
  - Last resort -- use when there is no other choice
  - Sharpen critical thinking

Parameter Estimation Error

- The current parameter estimates, say $\varphi\hat{}$, if used for all predictions, contribute bias to the observed prediction errors, $\{y^E - y^M\}$
- Therefore, $var_{\varphi\hat{}}(y^M)$ is not a contributor to the variance of the observed prediction errors
  - It is inappropriate to compare observed prediction errors to a variance that includes $var_{\varphi\hat{}}(y^M)$

Comment:
Measuring Predictive Capability *vis a vis* UQ

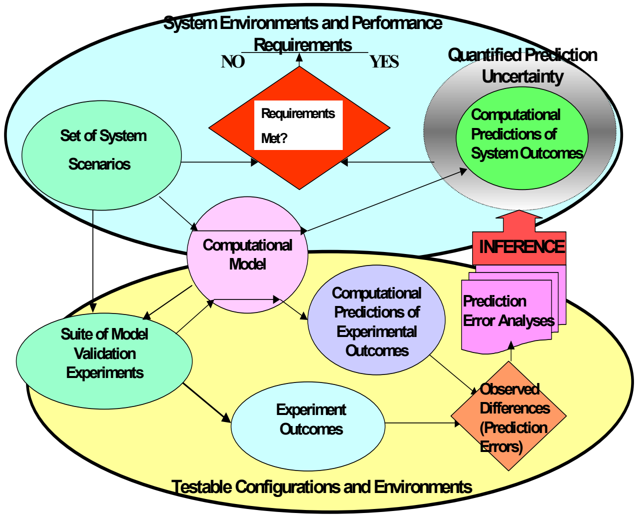
- As foam case study illustrates, predictive capability is measured via
  - analysis of $\{x, y, y^M\}$ data
  - no conventional UQ exercise on y^M was required
    - » *(except to evaluate effect of x measurement error)*

    (UQ = uncertainty quantification - generally the propagation of x or φ dist'ns. through M)
- ***The relevant prediction uncertainty is the difference between nature and model. UQ exercises on the model alone CANNOT tell you anything about nature vs. model.***
- UQ does have an important role: working problems that occur after predictive-capability is measured --
  - distribution prediction
  - merging results

Comments

You can't infer prediction error cloud by exercising the model

In case study, I didn't have to do any Monte Carlo sorts of analysis, in contrast to what some people claim.



Extension: Distribution Prediction

- Suppose x has an *assumed* probability distribution over some set of scenarios
- Problem is to predict resulting dist'n. of y

- Under the statistical model for y,

  $y_x = y_x^M + e_x; \quad e_x \sim (\beta_x, \sigma_x),$

  by the law of total variance:

$$\mathrm{var}_x(y_x) = \mathrm{var}_x(y_x^M) + E_x(\sigma_x^2) \quad \text{(when } \beta_x = 0\text{)}$$

- In words:

  nature's variance = model-based variance + extra-model variance

Comment

For this relationship:

$$var_x(y_x) \;=\; var_x(y_x^M) \;+\; E_x(\sigma_x^2) \qquad (\text{when } \beta_x = 0)$$

- Stochastic propagation techniques - estimate the first right hand term
- Model-Validation experiments and analyses - estimate the second right hand term
- Many "uncertainty" analysts work the first term; ignore the second *(and claim they're evaluating prediction uncertainty!)*, thereby underestimating variability, thereby overestimating reliability, …
- Both are needed for distributional predictions

UQ Issue: Variability vs. Estimation Uncertainty

- Generally:
  - x's: variables that could physically vary (depending on scenario of interest)
    - » *e.g., mission variables -- impact velocity and angles*
  - φ's: unknown constants, estimated with error
    - » *e.g., coefficients in equations of state.*
- Treating variability and estimation uncertainty probabilistically, then mixing them is really not interpretable -- apples and oranges.
- Some in probabilistic risk analysis community now separate treatment of x and φ:
  - nested Monte Carlos
  - illustrative result: with 90% "confidence" the probability of failure is between .005 and .017
    - » *(vs. the estimated probability of failure is .010).*

Issue: Validation as Hypothesis-Testing

- Some researchers treat model-validation as a hypothesis-testing problem:
  - Test: H_0: $E(e_x) = 0$
  - compare $\{y^E - y^M\}$ to constructed σ $\{= \sqrt{(\sigma_E^2 + \sigma_M^2)}\}$
- Even if hypothesis is not rejected, this does not mean e_x is negligible or can be ignored in characterizing predictions
- In fact, the noisier e_x is, the more likely it is that the model will 'pass' validation testing!

Model-validation is (should be) estimation, not hypothesis testing.

Issue:
Surely this approach requires too much testing!

Scientific assessment of predictive-capability probably requires more experimentation than envisioned by current methods

(vu-graph norm, ocular metric),

BUT

- The foam case study is model of higher-level testing and measurement of predictive-capability
  - » *focus on small no. of x-variables, linear regions*
  - » *small no. of tests*
- In some cases, we will be able to merge predictive-capability info from more numerous lower-level tests to derived measurement of predictive-capability at application level

Analysis Issue: Putting it all together

- Research Issue: How to combine prediction error data/models from different levels to infer prediction capability for application?
- One possibility:
  - $y_A^M = M(y_1, y_2, \ldots, y_k)$
  - $y_i^M = m_i(x_i : \varphi_i)$
  - $y_i = y_i^M + e_i$ (from predictive-capability expts. on m_i)
  - Analysis: propagate estimated e_i distributions through M; estimate resulting distribution of e_A and characterize precision of that estimate
- Example: Separate models for:

$$y_1 = \text{stress};\ y_2 = \text{strength}$$

Combined model:

$$y_A = \text{margin} = y_2 - y_1$$

Model-Confidence in the News

DoD comparison of computer simulations versus live fire tests of the effect of gunfire on helicopter blades:

- On a scale of 1 to 10, the models scored:
  - 7 in predicting how the shell would penetrate the blade,
  - 3 in predicting the destruction of the helicopter blade,
  - 2 in predicting the loss of a helicopter,
    - *[Sandia Daily News, 10/17/96]*
- modeling hierarchy: phenomenon - component - system
- predictive capability decreases as complexity increases
- validation scoring rule? ….?